RISK MANAGEMENT AND DATA QUALITY SELECTION: AN INFORMATION ECONOMICS APPROACH

Arun Bansal
Robert J. Kauffman

Department of Information, Operations, and Management Sciences
Leonard N. Stern School of Business, New York University
44 West 4th Street, New York, NY 10012

RISK MANAGEMENT AND DATA QUALITY SELECTION: AN INFORMATION ECONOMICS APPROACH

by

Arun Bansal
Leonard N. Stern School of Business
New York University
New York, NY 10003

and

Robert J. Kauffman
Leonard N. Stern School of Business
New York University
New York, NY 10003

November 1991

Center for Research on Information Systems
Information Systems Department
Leonard N. Stern School of Business
New York University

Working Paper Series

STERN IS-91-37

We wish to thank Robert Mark, Brian Fitzpatrick, Joonkee Hong, Jonathan Lomartire, Bruce Hannan of Manufacturers Hanover Trust Company for generously sharing their time and knowledge of risk management in the investment banking industry. We also appreciate the critical advice of Ted Stohr, Bruce Tuckman and Rob Weitz at the Stern School of Business at New York University.

Risk Management and Data Quality Selection: An Information Economics Approach

ABSTRACT

Data quality has been shown to be a major determinant of the value of systems that utilize input data feeds and transform them into valuable information under a variety of business contexts. For this study, we have chosen a financial risk management context to investigate the relationship between data quality and value of risk management forecasting systems. Three attributes of data quality, frequency, response time, and accuracy, along with the cost of data are considered. Joint impacts of attributes are also considered. It is shown that an increase in report frequency results in an increase in the utility of a risk management forecasting system, but this increase is limited by the responsiveness of the hedging scheme. Frequency is shown to improve the utility of the forecasting systems in two ways: *First*, an increase in frequency pushes the predicted states closer to the actual states and *second*, an increase in frequency causes the reliability of the forecasting model to increase. A delay in response time of reports is predicted to have a greater impact on utility for high frequency reports than for low frequency reports. Finally, data inaccuracies are recommended to be the first concern of a portfolio manager before an attempt is made to increase the reporting frequency.

1. INTRODUCTION

In the financial services industry, a significant amount of money is spent on information technology. A substantial portion of it is aimed at improving a firm's ability to make quick, effective and profitable decisions, to bring new products to market ahead of the competition, to continuously track any changes in the key market indicators, and to enable better control of business risks. In this context, *risk* is defined as "the lack of predictability of outcomes" affecting the set of financial transactions and positions which cumulatively form the firm's business [DOHE85, p.15]. Thus, risk includes the possibility of both pleasant surprises as well as adverse business outcomes. *Risk management* is the management of the resources and commitments of a firm so as to maximize its value, taking into account the unpredictable outcomes that can affect the firm's performance.

Consider the following examples of bad risk management and its consequences in last ten years or so:

- In 1986, 145 American banks failed or were merged with other organizations; another 1484 were officially reckoned to be in trouble at the end of the year [STEV87].

- The percentage of loans that American banks wrote off as uncollectible grew by 50% from 0.57% in 1982 to 0.86% of total loan assets by 1986 [STEV87].

- Svenska Handelsbank lost approximately 100 million dollars in options trading because it was trading derivatives without adequate control systems in place [SHAL89].

Learning from these and similar losses during the late 1980s, firms started building sophisticated systems to monitor and measure global risk. Such systems are normally referred to as "risk management systems." Chicago Research and Trading, for example, recently deployed a system for its dealers in New York. Other firms, such as Quotient Inc. and Devon System which specialize in risk management systems for interest rates derivatives and foreign currency options, have taken on the role of value-added vendors, developing risk management systems for resale to financial institutions. Meanwhile, other investment firms, including Merrill Lynch and Shearson Lehman, have set up special risk management units to monitor the risk by-product, by currency and by geographic region, and have also deployed automated systems to support risk management.

However, the cost of building such systems is very high. A risk management system today can cost from $10 million to $25 million in addition to several million dollars a year in maintenance fees [SCHM90B]. A portion of these expenditures involve periodic and recurring purchase of data. Large financial firms track the market by using video and digital feeds of real-time data. *Video data feeds* contain the video images of fixed format pages of data about the market or a group of financial instruments. *Digital data feeds*, on the other hand, contain digital data which can be unbundled and used for value-added analytics. Examples of electronic data sources in this area include the Chicago Board Options Exchange, New York Stock Exchange ticker, and market-specific wire services from firms such as Reuters, Telerate, Market Vision, Knight-Ridder and Dow Jones [AREN89, SCHM90A]. These secondary quote vendor services also consolidate information from different exchanges for page, graphics or ticker-formatted transmission and presentation.

As quality of the data improves, so should the quality of decisions that are made using this data. The selection of data vendor services involves a trade-off between the quality of decisions made and the cost of attaining this quality. For example, infrequent market indicator updates are less expensive than frequent updates. In absence of any formal guidelines it is difficult to estimate exactly what frequency is appropriate. This situation is further complicated by the presence of other dimensions of data quality, including frequency, response time and accuracy. The nature of the interaction of these dimensions becomes a major determinant in the selection of appropriate data quality.

1.1. Research Questions

While there are open questions for IS researchers and practitioners regarding the impact of data quality on decision making in the context of risk management, our premise is that making progress on this problem will require decomposing a risk management system into two separate components: the model employed to arrive at risk management decisions, and the data that describe changes in the states of the world that represent risks for the firm. Each of these two components can be considered separately and their impacts on performance evaluated. We call these components collectively as *risk management technology* **(RMT)**. In the present study, we focus on the data component of RMT. By concentrating on the quality of digital data feeds, we are taking the first step towards measuring the potential for RMT to deliver business value to the firm. There are three data quality attributes that are considered to be important by senior management for assessing financial risk: frequency, response time, and accuracy. Together with the decision making benefits of these data quality attributes, we must also consider the data acquisition cost for a complete cost-benefit analysis. How these factors affect the quality of risk management decisions is the focus of the present study.

Our research questions are as follows:

- What is the impact of data quality on a risk management system's assessment of risk?
- How does data acquisition cost affect management's decisions regarding appropriate digital feed configurations across various risk management contexts?
- How frequently should information be obtained for assessing risk accurately?
- How does the interaction of data quality attributes affect management selection

of an *optimal* data feed configuration?

In trying to answer these questions, we hope to guide future efforts by information systems researchers in the area of risk management. We will also illustrate how the data component of RMT can be investigated in terms of its cost efficiency. Finally, the conceptual framework we use for this study can be used as a basis for future empirical studies.

1.2. Outline of the Paper

In the remainder of this paper we develop a theoretical foundation to answer the questions posed above. Section 2 reviews concepts from risk management, information economics, management science and accounting models of information quality that are required to build a foundation for the approach we propose. Section 3 investigates the impact of relevant data quality attributes on financial risk. Section 4 develops a number of propositions that address the individual and joint impacts of three data quality attributes on performance of a financial risk management forecasting system. The propositions show how optimal data quality will differ depending on the relationship between utility and cost curves.

2. PRIOR RESEARCH

A number of studies examined data quality and its impact on the value of information systems [HILT81, DEMS85, BARU89, AHIT89]. However, there are two differences between the present study and the studies that are already available in literature. ***First***, in this study we have broken up an information system (RMT in this case) into its two components: a data component and a model component and we focus only on the former. This enables us to investigate the impacts of data quality on performance of information systems that do not depend on the choice of the model selected for the system. ***Second***, in contrast to earlier studies, the present study investigates both the individual impact of a data quality attribute and the joint impact of several data quality attributes on performance of an information system. We define *individual impacts* of data quality attributes as those where only one input data quality attribute is assumed to vary and to cause variation in the economic value of an information system. *Joint impacts*, on the other hand, are defined as those impacts on the value of an information system that involve the simultaneous variation of more than one data quality attribute.

Prior research efforts to investigate data quality impacts have been inter-disciplinary. Information economists, accounting researchers, management scientists and risk managers are all interested in knowing more about the impacts of data quality on the performance of their systems. To integrate the available studies around a common theme, we evaluate the major findings in each of these disciplines as they relate to the impacts of data quality.

2.1. Risk Management, Information Economics and Data Quality

Researchers have looked into a variety of data quality attributes to understand how data quality affects the utility of an information system. According to Ahituv [AHIT89], these attributes can be appropriately divided into four categories: timeliness, contents, format and cost [AHIT89]. In the timeliness category, attributes such as frequency, recency and response time are included. In terms of contents, accuracy, relevance, aggregation level and exhaustiveness are considered. Finally, format includes the medium, color, presentation strategy and sequence of presentation of data.

For risk management, frequency, response time and accuracy are considered important quality dimensions which play a crucial role in economic evaluation of data feeds[1]. *Frequency* determines the interval of time between successive reports about the current state of the world. *Response time* measures the time taken to report a certain value for a particular state of the world. Data exhibits *accuracy* when the actual state of the world it measures matches the value reported by the measure.

[1]Although qualitative measures of the value of an IS (e.g., user satisfaction) are important, they are not the focus of this study. We concentrate instead on economic aspects of the value of information systems.

The primary studies that we will discuss below are shown in Figure 1.

INSERT FIGURE 1 ABOUT HERE

2.2. Information Economics and Data Quality Estimation

The tradeoffs involved in selection of data feeds of appropriate quality can be characterized from an economics perspective in terms of the value of information they provide. *Value of information* is defined as the difference between the utility derived from using information and the cost of obtaining it [DEMS85, HILT79, HILT81, AHIT89]. The value of an information system has been shown to depend on three primary factors [HILT79, DEMS85]:

- the signals generated by it;
- the accuracy by which these signals predict a certain state of the world;
- the actions that are taken to maximize the payoffs corresponding to these states.

Prior work in this work has evaluated models with both continuous data quality variables [HILT79, HILT81] and discrete data quality variables [DEMS85]. According to Demski, for example, managers in the business world generally have limited states of the world to consider, they understand only limited types of signals and they have a limited number of actions available to respond to the indicated states of world. Hence, a discrete distribution of states of the world and signals generated to measure these states should suffice to estimate the business value of information generated by an information system.

Other studies in this area have also investigated the relationship between one or more system attributes and the value of information produced by a system [BLAC53, IJIR73, WILS75, GROS80, McCA85, KIHL74]. Although most of these studies have evaluated the individual impacts of data quality attributes on the value of information in a variety of decision making contexts, only a few have examined the joint effects of these attributes [BARU89]. While individual impacts are important for the evaluation of data feeds, joint effects also need to be examined more closely. For example, the impacts of accuracy and response time on the value of data feeds may be different for reports of low or high frequencies. The present literature in information economics has very little to offer on the joint effects of these data quality attributes.

2.3. The Value of Data Quality from the Accounting and Management Perspectives

Other areas of business, including accounting and management science, have considered the question of evaluating data quality for decision making [CUSH74, BAGM75, BALL85]. Designing internal control systems, defining which data items to track for control purposes and recommending the length of time between successive management and auditing reports are representative of the areas studied. For example, Cushing [CUSH74] studied internal control systems used by controllers and auditors to identify the parameters that are relevant for ensuring control. Internal control systems are comprised of one or more procedures that are directed at detecting errors. The problem can be viewed in terms of the propagation effects of errors through a hierarchy. Cushing studied the effects of controls on process probabilities, timing of control and the effect of transaction size on system probabilities, among other issues. However, he ignored the concept of utility and economic value in an operational or business process, leaving it as a subject for future research and, although individual effects were investigated in Cushing's study, joint impacts of transaction size, process probabilities and timing of control on value of internal control systems were not explored.

Bagman [BAGM75] addressed the issue of evaluation of internal information systems within a multi-person world. One of the contributions of his work was to show that in multi-person competition, a "coarser" information system will be more valuable to an individual than a "finer" one. Broadly speaking, finer information is more detailed and less aggregated than coarser information. The conclusions of the study could thus be used by senior management to evaluate investments in obtaining finer information.

Ballou and Pazer [BALL85] addressed the accounting concerns of data quality and provided a formal management science model for carrying out sensitivity analysis. Moreover, their model fits the problem of gauging data quality in risk management application quite well. The framework assesses the impact of data and process quality upon multi-input and multi-output information systems. The authors employ the concept of a *loss function* to study these impacts. A loss function characterizes the effect of an error, Δx, in an input data variable x, on the output, y given by the transformation function, F(x). The approach adopted by Ballou and Pazer is remarkably similar to one proposed by Hilton [HILT81] and Demski [DEMS85]. While Hilton and Demski considered the individual benefits of using a certain level of information quality, Ballou and Pazer used the *opportunity cost* of not having high quality data. They only implicitly consider the joint impacts of data quality attributes because they treat the overall effect of several data quality attributes on the system output through a continuous and double differentiable loss function.

Ballou and Tayi [BALL89] use an integer programming model to allocate a firm's resources for data quality enhancement. Their model uses direct savings that result from applying a data quality improvement procedure as the benefits of improving data quality. They assume that managers work under a constrained budget for such data quality improvements. Hence the model they propose deals with the joint effects of data quality attributes and the cost of making improvements to them.

In summary, a lot of work on evaluation of data quality attributes has been accomplished in a number of functional areas of business research. But unless an attempt is made to synthesize the findings of previous studies to address the joint impacts of multiple data quality attributes, the findings can not be fully exploited for a variety of IS evaluation situations.

3. DATA QUALITY ATTRIBUTES AND RISK MANAGEMENT

To illustrate how information economics can be applied to optimize data quality, we next present an information economics model that applies to a financial risk management domain. We also illustrate how the findings of our model can be used in practical risk management situations by providing a data quality optimization example involving the prediction of prepayment rates for mortgage-backed securities.

3.1. An Information Economics Model for Risk Management

For the discussion that follows, we will only consider risk management situations where forecasting models are used to predict states of the world. Our purpose is to investigate the decisions that are under management's control involving the relationship between data quality and the performance of a risk management forecasting system (**RMFS**).

Generally a set of signals y_1 to y_n, given by $\Re(y_1,\ldots,y_n)$, is used in a forecasting model to predict the states of the world s_p for a financial instrument. These signals may be available at different frequencies, can have varying response times and can possess varying degrees of inaccuracy. The predicted states can be transformed to provide a measure of risk in investments in a set of securities. To diversify this risk, a portfolio manager "hedges" his investments by going long (acquiring) in some and short (selling) in others. Hedging offsets undesirable risk by creating an invested position in an instrument that moves inversely with the risky position. The goodness of a hedge is defined in terms of its *hedge efficiency*, the ratio of returns that would result from the hedge to the maximum returns possible for same specified level of risk [WEST86]. Since, the hedge efficiency is a function of how close the predicted state is to the actual state, s_a, the utility of the RMFS will depend on the proximity of the predicted state to the actual state when it is used to create the hedge.

In the absence of a forecasting model a portfolio manager must guess the financial states of different investments in the future. Let the guessed state be given by s_g. Then, the utility of an RMFS can be calculated using a model that is based on earlier models by Hilton [HILT81] and Demski [DEMS85], and recast for a forecasting context:

$$U(h)_{fr\varepsilon} = \int_{\Re_h \in \mathbb{R}_h} \max_{\alpha \in A} \int_{s_a, s_p \in S} u[\eta_{hedge}(|s_a - s_p|, \alpha)]\, p(|s_a - s_p| \mid \Re_h(y_1, \ldots, y_n)_{fr\varepsilon}, m_h)\, p(\Re_h(y_1, \ldots, y_n)_{fr\varepsilon}) - \max_{\alpha \in A} \int_{s_a, s_g \in S} u[\eta_{hedge}(|s_a - s_g|, \alpha)]\, p(|s_a - s_g|) \quad \forall f \in F \;\; \forall r \in D \;\; \forall \varepsilon \in E \qquad (1)$$

The variables in the model are defined as follows:

$\Re(y_1,\ldots,y_n)$ — the set of n signals used by risk management forecasting system h, that is used to predict the future states of investments;

$\mathbb{R}$ — the set of signal sets, $\Re(y_1,\ldots,y_n)$, that represent the range of possible signals that can be reported by forecasting system h;

$U(h)_{fr\varepsilon}$ — the utility of risk management forecasting system h, when a design choice is made with respect to three data quality attributes of the signal set $\Re$:

- f, the frequency, in the set of all possible frequencies, F, of the signals,
- r, the response time, in the set of all possible response times, D, that the signals can have,
- ε, the extent of the inaccuracies, in the set of all possible errors, E, that can occur in the signals;

S — the states of nature, including --

- s_p, the state predicted from a given signal set, $\Re$,
- s_g, the state guessed in the absence of forecasting system h, and,
- s_a, the actual state that occurs;

A — the set of all possible financial instruments, α, that can be used to hedge the instrument whose future states are being predicted by forecasting system h;

hedge — an investment in a specific financial instrument, α, that is effected to control the risk of an invested position;

η_{hedge} — the efficiency of the hedge created between the invested position and the financial instrument selected for the hedge, α;

m_h — the model used by risk management forecasting system h;

$p(\Re(y_1,\ldots,y_n))$ — the prior distribution of signals from forecasting system h;

$p(\mid s_a - s_p \mid \mid \Re(y_1,\ldots,y_n)_{fre}, m_h)$
the posterior distribution of the absolute difference between the actual state, s_a, and the predicted state, s_p, of the investment position;

$p(\mid s_a - s_g \mid)$ — the prior probability of the absolute difference between the guessed state, s_g, and actual state, s_a, of the investment position.

In this model, the utility of a risk management forecasting system, h, with the data quality characteristics, frequency (f), response time (r) and accuracy (ε) for a specific investment position has two major components. The first term

$$\int_{\Re_h \in R_h} \max_{\alpha \in A} \int_{s_a, s_p \in S} u[\eta_{hedge}(|s_a - s_p|, \alpha)]\, p(|s_a - s_p| \mid \Re_h(y_1,\ldots,y_n)_{fre}, m_h)\, p(\Re_h(y_1,\ldots,y_n)_{fre}) \tag{2}$$

gives the maximum utility that can be derived from the forecasting model by selecting the best possible hedge on the basis of information obtained from the system. This component of utility is also dependent on the reliability of the forecasting model and the prior distribution of the signals.

The second term

$$\max_{\alpha\in A} \int_{s_a, s_g \in S} u[\eta_{hedge}(|s_a - s_g|, \alpha)] p(|s_a - s_g|) \tag{3}$$

gives the maximum utility that can be derived by purely guessing the future states of nature for the current investments.

Four determinants of this utility are identified as:

- the structure of the choice set, given by different possible hedges that can be made to spread risk;
- the structure of the payoff function which maps the hedge efficiency and alternative hedging options into a measure of payoff; (This efficiency is also dependent on the availability of different financial instruments that can be used to create the hedge)
- the predictive ability of the forecasting model that maps input signals into a predicted state;
- the degree of uncertainty in the input signals.

Thus, overall the model describes the difference in the utility associated with the hedge efficiency that results when the appropriate hedge for the investment position is determined using the signal set produced by the risk management forecasting system or by simply guessing about future states of the investment position. Since the model considers the possibility of many potential signal sets and also many possible hedges, we can see that variations in data quality can play an important role in determining utility.

3.2. Risk Management for Mortgage-Backed Securities: Background

The mortgage industry has two markets: a primary market and a secondary market [PINK87]. Lending institutions directly deal with the customers (or borrowers) in the primary market. Borrowers in the primary market are obliged to pay the principal and the interest to the lender. The lender, in turn, is interested in raising capital and spreading the risk of making such loans with the help of federal agencies. The federal agencies purchase a bank mortgage portfolio or individual loans through a process called "securitization." The securities created in this way are called "mortgage-backed securities" (**MBS**). These securities trade freely in the secondary market just like any other financial instrument. The lenders or the servicing agencies collect payments from customers and, after deducting their servicing fees, pass them to the current holder of the MBS created to securitize the loan [HAYR89].

MBS are considered to be fixed income investments. However, the borrower retains an option to prepay the loan at any time and this makes investments in MBSs risky. Prepayments act as a *call option* on a fixed income security adding uncertainty during the period of investment. In the case of rising interest rates, the holder of an MBS can expect good returns if prepayment occurs; the holder can invest prepayment in another security at a higher interest rate. On the other hand, prepayments occurring in a time of falling interest rates will lead to a loss. We define the *prepayment rate* as the percentage of total customers that prepay a loan at that specific point of time. Thus, whatever the case, a knowledge of the prepayment rate and an ability to predict future interest rates are essential for assessing risk [PINK87].

Currently, forecasts of future prepayment rates are made using regression models [HAYR89]. These models estimate future prepayments based on current and past values of several relevant variables. These include: the age of the mortgage; the difference between the coupon rate of the mortgage and the current interest rate; and macroeconomic indicators, such as individual well-being, consumer confidence and GNP. The financial industry needs information on each of these indicators to forecast prepayments in the future to reduce the risk of managing portfolios of mortgage-backed securities. However, the question of determining *optimal data quality* (e.g., the interval between two successive reports, the delay in receiving reports and permissible inaccuracies) is important as recurring data feed requirements can be very expensive.

3.3. Impact of Data Quality on the Value of an RMFS

Although the frequency and accuracy of reports exhibit a direct relationship with the utility of an RMFS, response time generally bears an inverse relationship. Thus, if frequency or accuracy is high, the utility of an RMFS is also expected to be high. On the other hand, if response time is high, the utility of information is likely to be low. This is especially true in this domain of risk management, where new information has value only if it is received on time; delayed information usually limits the number of actions that can be taken by a risk manager.

While individual impacts of data quality attributes on the utility of an RMFS are easier to study, the joint effects of two or more such data quality attributes are often difficult to predict and thus are poorly understood. For example, response time exhibits a varying and complex relationship with other data quality attributes, such as report frequency. For reports that have the same response time, the ratio of decline in utility to the original utility is higher for reports with high frequency than for reports with low frequency. We will deepen this discussion with a series of propositions and proofs in the next section.

With variations in data quality, the predictive ability of the forecasting model is affected, thus affecting the utility of the risk management forecasting system. Compromises in data quality normally are associated with a lower cost to obtain it. The concept of the value of information best incorporates the trade-off between the utility of risk management forecasting information and the cost of producing it, as given by the following expression:

$$V(h)_{fr\varepsilon} = U(h)_{fr\varepsilon} - C(h)_{fr\varepsilon} \tag{4}$$

where

$V(h)_{fr\varepsilon}$ = the value of risk management forecasting system h with report frequency f, response time r and inaccuracy level ε.

$C(h)_{fr\varepsilon}$ = the cost of the system h.

4. COST-BENEFIT ANALYSIS FOR DATA QUALITY IN RISK MANAGEMENT

A forecasting model should increase the utility of an RMFS with an improvement in data quality in at least two ways. *First*, with an improvement in data quality the predicted states are likely to be closer to the actual states, and hence the efficiency of the hedge selected for the security whose states are being predicted should improve. *Second*, with an improvement in data quality, the reliability of the forecasting model improves, further improving the utility of the RMFS.

We use the model proposed in the prior section to investigate how frequency variation affects the utility of an RMFS individually and jointly with variations in other attributes, such as response time and inaccuracy in reports. We also investigate the optimal report frequencies for an RMFS based on a cost-benefit analysis. We state and prove six propositions which illustrate the relationship between the data quality attributes and the utility of an RMFS.

The following assumptions will be used in testing our propositions:

(1) Only one risk management forecasting model, m_h, is used for the RMFS h.

(2) Data feeds exhibit variation in quality for a given risk management scenario.

(3) A rational risk manager is a value maximizer who selects a financial instrument, α, to maximize hedge efficiency for an investment position based on the knowledge of the predicted future states of the security being hedged.

(4) A risk manager's utility linearly increases with an increase in hedge efficiency.

(5) Data feeds are available only in integer multiple frequencies (i.e., the ratio of a higher report frequency to a lower one for the same data feed is an integer).

4.1. Reporting Frequency and Utility of a Risk Management Forecasting System

Predictions from an RMFS about the future state of an MBS are used to select other securities with which to create the hedge. If the predictions are inaccurate, the hedges formed using these predicted states will not be efficient. Thus, the hedge efficiency can be used as a measure to gauge the affects of data quality variations on utility of the forecasting system. We characterize a hedging scheme in terms of its degree of responsiveness, as defined below:

Definition 1: The *responsiveness of a hedging scheme* is defined as the ease with which alternative financial instruments can be found to hedge a given financial instrument.

The degree of responsiveness of the hedging scheme affects the rate of variations in hedge efficiency. A hedging scheme is considered to have a *high responsiveness* to changes in predicted states if the hedge efficiency varies linearly with changes in the predicted states. On the other hand, a hedging scheme has *low responsiveness* if the variations in predicted states do not affect the hedge efficiency at all.

When a hedging scheme is highly responsive, an increase in reporting frequency will always be beneficial as shown by the following proposition.

Proposition 1: For a hedging scheme that has high responsiveness, the utility of an RMFS that predicts future states of the world, s_p, for a financial instrument included in the hedge is non-decreasing with an increase in reporting frequency given that the reporting frequencies are integer multiples of each other and other data quality attributes remain unchanged.

Proof: Given that the hedging scheme has high responsiveness, it is always possible to find a financial instrument that can act as a hedge for the security whose future states are to be predicted. With a highly responsive hedging scheme, η_{hedge} should vary linearly with variations in the absolute difference between the predicted and the actual state, given by $|s_a - s_p|$ and should increase with a decrease in the value of $|s_a - s_p|$ and vice-versa.

To prove that the utility of a RMFS is non-decreasing in frequency f, we need to differentiate the expressions inside the integral sign of the first term in Equation 1 (the utility estimate of the hedge selected in the presence of risk management forecasting information) with respect to f and show that the result is non-negative.

We rewrite the expression inside the integral sign as follows:

$$T = u(\eta(\Delta s,\alpha))p(\Delta s|R,m_k)p(R) \tag{5}$$

where η is η_{hedge}, Δs is $|s_a - s_p|$, and R is $\Re(y_1 \ldots y_n)_{fre}$. Now differentiating T with respect to f yields:

$$\frac{\partial T}{\partial f} = \frac{\partial u(\eta(\Delta s,\alpha))}{\partial f}p(\Delta s|R,m_k)p(R) + u(\eta(\Delta s,\alpha))\frac{\partial p(\Delta s|R,m_k)}{\partial f}p(R) + u(\eta(\Delta s,\alpha))\frac{\partial p(R)}{\partial f}p(\Delta s|R,m_k) \tag{6}$$

This equation has three components. The *first* gives the variation in utility due to the change in predicted state when the frequency of data feeds is varied. The *second* component denotes the variation in reliability of forecasts when the frequency is varied. The *third* component gives the variation in prior distribution of signals with variations in reporting frequency.

Note that $\partial u/\partial f$ can be decomposed into $(\partial u/\partial \eta)\ (\partial \eta/\partial(\Delta s))\ (\partial(\Delta s)/\partial f)$. Given our assumption that utility is linear and increasing in hedge efficiency η (Assumption #4), in the above expression $\partial u/\partial \eta \geq 0$. Also, hedge efficiency η should decrease with an increase in Δs since with an increase in Δs, the hedge designed on predicted state s_p will not be as efficient as the hedge that could have been created had the actual state s_a been known. Thus, it follows that $\partial \eta/\partial(\Delta s) \leq 0$. Now, since we expect the predicted state s_p to be closer to the actual state s_a with an increase in frequency, we have $\partial(\Delta s)/\partial f \leq 0$. Using these results, it follows that $\partial u(\eta(\Delta s,\alpha))/\partial f \geq 0$. Since the other two terms of the first component in (6) are probabilities, they should be greater than or equal to zero. Hence, it follows that the first component $\partial T/\partial f$ is non-negative.

Since the hedge efficiency, η, is between 0 and 1 and utility u is linear in η, it should be non-negative. With an increase in reporting frequency, there should also be an increase in the reliability of the forecasting model, resulting in $\partial p(\Delta s/R)/\partial f \geq 0$. Thus, the second component of (6) is also non-negative.

With randomly generated signals, the probability of occurrence of a signal set R should not vary with a variation in reporting frequency, and thus we know that $\partial p(R)/\partial f = 0$. Thus, the third component of (6) is zero. This enables us to conclude that $\partial T/\partial f \geq 0$, completing the proof.▪

Discussion: This proposition shows that higher report frequency increases the utility of a forecasting system in two ways: by pushing the predicted states closer to the actual states and by increasing the reliability of the RMFS. Hence, if cost were not a criterion, to maximize the utility of an RMFS a value maximizing portfolio manager would always attempt to use the highest possible frequency data feeds.

Example: To illustrate the usefulness of this proposition, we will consider an example involving mortgage-backed securities in more detail. Assume that prepayment rate of a mortgage-backed security is predicted based on a set of observations from the past. A number of indicators are used to predict these rates and a regression model is used to make such predictions [HAYR89]. Depending on the predicted prepayment rate, a suitable hedge with other securities can be formed to keep the risk in the portfolio of the two securities under a pre-specified level. For this example, we assume that utility to the portfolio manager from using the system is a linear transform of hedge efficiency η and is given by $k_2\eta$.

We also assume that there are a number of alternative instruments available to hedge the MBS. Thus, even with small variations in predicted prepayment rates, the hedges and the hedge efficiencies should vary. We assume that the hedge efficiency is realized as a constant multiplied by the extent of the overlap (1- the absolute deviation % (Δ)) in actual and expected prepayment rates. Mathematically, η is given by the relation $k_1(1-\Delta s)$, where k_1 is a constant. Note that the hedge efficiency lies between 0 and 1 as Δs is always less than or equal to 1.

We further assume that Δs varies with frequency f as given below. But, note that although Δs varies with response time, r, and report inaccuracy, ε, we assume that both r and ε remain unchanged for this proposition.

$$\Delta s - k_3 - k_4 f^2 r^{-2} \varepsilon^{-2}. \tag{7}$$

The expression above shows that higher frequencies reduce the deviation, higher response times amplify the deviation and higher error rates in the data amplify the deviation. From the boundary condition that Δs is likely to reach a maximum, Δs_{max}, when the frequency f equals zero, and by taking into account that Δs is non-negative, it follows that k_3 equals

Δs_{max}.

Combining these assumptions, the expression for $u(\eta(\Delta s,\alpha))$ can be written for this example as:

$$u = k_1k_2 - k_1k_2\Delta s_{max} + k_1k_2k_4f^2r^{-2}\varepsilon^{-2}. \qquad (8)$$

Reliability of the model is assumed to be increasing in f, r and ε and the increase is given by the following relationship.

$$p(\Delta s|R,m_h) = k_5f^2r^{-2}\varepsilon^{-2}. \qquad (9)$$

Now the relationship between the utility of the RMFS, $U(h)_{fr\varepsilon}$, and the frequency f can be shown by the terms inside the integral as follows.

$$T = k_1k_2k_5(1-\Delta s_{max})f^2r^{-2}\varepsilon^{-2} + k_1k_2k_4k_5f^4r^{-4}\varepsilon^{-4}. \qquad (10)$$

where coefficients k_1, k_2, k_4, k_5 are greater than or equal to zero. For expository convenience, we further assume that constants k_1, k_4 and k_5 have the value 1. This reduces (10) to:

$$T = k_2(1-\Delta s_{max})f^2r^{-2}\varepsilon^{-2} + k_2f^4r^{-4}\varepsilon^{-4}. \qquad (11)$$

Clearly,

$$\frac{\partial T}{\partial f} \geq 0. \qquad (12)$$

Although we have shown that an increase in frequency will not lead to a decrease in utility if some of the assumptions mentioned above hold in a specific risk management scenario, a portfolio manager would also be interested in knowing how the rate of change in utility varies with a change in report frequency -- whether the utility of the RMFS is convex, concave or linear in f.

This is important to determine optimal report frequency, as we will soon show. A portfolio manager may not be concerned about improving the report frequency of data feeds if the *marginal increase* in utility of the RMFS is not going to be significant for any such improvements. Our next proposition presents one scenario where improvements in reporting frequency will provide high returns.

Proposition 2: For a highly responsive hedging scheme, the utility of an RMFS will be convex increasing with an increase in report frequency f, when both the deviation $|s_a - s_p|$ is concave decreasing and the reliability $p(|s_a - s_p| \,|\, R,m)$ is convex increasing in f and other data quality attributes are invariant.

Proof: To show that utility is convex-increasing in f, we need to differentiate T twice with respect to f and show that $\partial^2 T/\partial f^2 \geq 0$ for the given conditions. This second derivative is given by:

$$\frac{\partial^2 T}{\partial f^2} = \frac{\partial^2 u}{\partial f^2} p(\Delta s|R,m_h)p(R) + 2\frac{\partial u}{\partial f}\frac{\partial p(\Delta s|R,m_h)}{\partial f}p(R) + 2\frac{\partial u}{\partial f}p(\Delta s|R,m_h)\frac{\partial p(R)}{\partial f}$$
$$+ u\frac{\partial^2 p(\Delta s|R,m_h)}{\partial f^2}p(R) + u\frac{\partial^2 p(R)}{\partial f^2}p(\Delta s|R,m_h) + 2u\frac{\partial p(R)}{\partial f}\frac{\partial p(\Delta s|R,m_h)}{\partial f}. \quad (13)$$

Since the distribution of signals does not depend on the report frequency, probability p(R) will be invariant with changes in f. Hence,

$$\frac{\partial p(R)}{\partial f}, \frac{\partial^2 p(R)}{\partial f^2} = 0. \tag{14}$$

This suggests that the third, the fifth and the last terms in $\partial^2 T/\partial f^2$ are equal to zero. In Proposition 1, we showed that utility of an RMFS and the reliability of its predictions both increase with frequency. Thus, the second term in $\partial^2 T/\partial f^2$ will always be greater than or equal to zero. For the fourth term to be greater than or equal to zero, the following should be true:

$$\frac{\partial^2 p(\Delta s|R,m_h)}{\partial f^2} \geq 0. \tag{15}$$

But we know this to be true because reliability is convex-increasing in f. Thus, to prove proposition (2) we only need to show that $\partial^2 u/\partial f^2 \geq 0$. Expanding $\partial^2 u/\partial f^2$, we get:

$$\frac{\partial^2 u}{\partial f^2} = \frac{\partial^2 u}{\partial f \partial \eta} \frac{\partial \eta}{\partial(\Delta s)} \frac{\partial(\Delta s)}{\partial f} + \frac{\partial u}{\partial \eta} \frac{\partial^2 \eta}{\partial f \partial(\Delta s)} \frac{\partial(\Delta s)}{\partial f} + \frac{\partial u}{\partial \eta} \frac{\partial \eta}{\partial(\Delta s)} \frac{\partial^2(\Delta s)}{\partial f^2}. \tag{16}$$

Since utility is assumed to be linear only with hedge efficiency η, the first term in (16) equals zero. Given a highly responsive hedging scheme, we can assume that the hedge efficiency is only linear in predictive accuracy, thus making the second term in (16) also equal to zero. Finally, from our earlier discussion in Proposition 1, utility is expected to increase with an increase in hedge efficiency η and η is expected to decrease with an increase in the predictive inaccuracy (as given by Δs) of the risk management forecasting system. Thus, for the third term in (16) to be greater than or equal to zero, $\partial^2(\Delta s)/\partial f^2$ should be less than or equal to zero. Since we know this to be true, it is clear that $\partial^2 p(\Delta s/R,m_h)/\partial f^2 \geq 0$ must hold. This completes the proof.■

Discussion: The above proposition shows that if predictive inaccuracy increases at an increasing rate with a decrease in frequency, and if the reliability of an RMFS increases at an increasing rate with an increase in frequency, then the utility of the RMFS is bound to increase at an increasing rate, provided the assumptions of the proposition are met. Since the utility is a direct measure of the payoff from a risk management forecasting system, a portfolio manager would obtain high returns by using highest possible data feed frequency if the predictive accuracy and reliability vary with frequency as given in our example.

Example: We again explore the MBS example presented earlier. Note that given the assumptions of the example, it follows that the accuracy with which prepayment rates can be predictive decreases at an increasing rate with an increase in report frequency as shown by the following equation:

$$\frac{\partial^2(\Delta s)}{\partial f^2} = -2r^{-2}\varepsilon^{-2} \leq 0. \tag{17}$$

Also note that the reliability of the forecasting model increases in frequency f with an increasing rate, i.e.,

$$\frac{\partial^2 p(\Delta s | R, m_h)}{\partial f^2} = 2r^{-2}\varepsilon^{-2} \geq 0. \tag{18}$$

Given these two conditions, we see that utility of the RMFS, $U(h)_{fr\varepsilon}$, rises at an increasing rate with a rise in reporting frequency as shown below:

$$\frac{\partial^2 T}{\partial f^2} = 2k_2(1-\Delta s_{max})r^{-2}\varepsilon^{-2} + 12k_2 f^2 r^{-4}\varepsilon^{-4} \geq 0. \tag{19}$$

In our discussion of Proposition 1, we saw that a highly responsive hedging scheme plays a role in determining how the utility of an RMFS varies with an increase in frequency. However, we did not discuss how these variations in utility would be affected if the responsiveness of the hedging scheme were low. If the decrease in responsiveness of a hedging scheme diminishes the effect of variations in frequency, it may not be economically beneficial for a portfolio manager to purchase digital feed data with a high reporting frequency.

Our next proposition shows how variations in the responsiveness of the hedging scheme affect the impact of frequency variations on utility of an RMFS.

Proposition 3: The increase in utility of an RMFS with a fixed increase in frequency will be greater for a more responsive hedging scheme than for a less responsive hedging scheme.

Proof: Let $\eta_1(\Delta s, \alpha)$ and $\eta_2(\Delta s, \alpha)$ represent two hedging schemes that use different financial instruments α depending on the predicted value of the state s_p of the instrument being hedged. η_1 is the more responsive of the two schemes to variations in predicted values of s_p. Let f_1 and f_2 be two frequencies at which reports indicating a particular state of the world become available. Assume that f_2 is greater than f_1. Let the state of the world at a certain point in time be shown as s_{p1} by reports of frequency f_1 and s_{p2} by the reports having a frequency f_2. As was shown in Proposition 1, lower frequencies will lead to a greater departure of predicted state from reality, increasing the value $| s_a - s_p |$.

Next, let the choice of an efficient hedge corresponding to s_a include α_1^* for hedging scheme η_1 and α_2^* for η_2. It follows from our assumption about responsiveness that the hedging efficiency will be higher for scheme η_1 than for η_2 if the true states of the world are not predicted by the reports. Also, since $| s_a - s_{p1} |$ is larger than $| s_a - s_{p2} |$, it follows that the difference in hedge efficiencies will be greater for s_{p1} than for s_{p2}. Thus, we can deduce that the reduction in utility depends on the responsiveness of the hedging scheme η.■

Discussion: This proposition shows that responsiveness of mapping functions may be a key determinant in selection of frequencies for reports. Thus, if the actions are not very responsive to key market indicators in risk management, then a minor increase in the accuracy of predicted states from an RMFS by using a higher reporting frequency is not going to increase the utility of the RMFS significantly.

Example: We again employ an example involving investments in MBSs to illustrate the intuition behind Proposition 3. In the example that illustrated Proposition 1, we considered a highly responsive hedging scheme, where a drop in predictive accuracy from an RMFS produced a comparable reduction in the hedge efficiency. Now, consider a slightly less responsive hedging scheme as given below:

$$\eta_{low} = .1(10(1-\Delta s))^{\frac{1}{2}}. \tag{20}$$

The more responsive hedging scheme, as shown earlier, is described by the following hedging efficiency calculation scheme.

$$\eta_{high} = 1-\Delta s. \tag{21}$$

The first scheme is less responsive than the second one because for the same amount of predictive accuracy of the RMFS (given by 1-Δs), the difficulty in finding a suitable hedging instrument in the first scheme results in a lower hedge efficiency in comparison to the second hedging scheme. Note that although the hedge efficiency is a function of Δs and α, the ease of finding α can be measured by the relationship between Δs and the hedge efficiency. We will illustrate this by an example. Suppose that the predicted rate and actual rates are off by 10%, i.e., 0.1. Then in the more responsive hedging scheme, it will be possible to find a suitable hedge instrument α and thus the hedge efficiency will be theoretically equal to 0.9. However, in a less responsive hedging scheme it will be difficult to find a suitable financial instrument, and the hedge efficiency (given by (20)) will be 0.3. Now consider a situation in which the changes in the financial markets lead to a change in prepayment rates. If the new rates are diverge from the original predictions by a value of 0.2, then the current hedge in the more responsive case is less desirable as the efficiency will fall to 0.8, a decrease of 0.1 from its original position. However, the efficiency of the less responsive hedging scheme will only fall by 0.017 to 0.283.

It should be clear from (20) and (21) that for the same loss in predictive accuracy due to a decrease in reporting frequency, a more responsive hedging scheme would produce greater reductions in the expected utility of the RMFS than a less responsive hedging scheme.

4.1.2. Joint Impact of Reporting Frequency and Cost of Acquisition and Handling the Data

Data acquisition cost is also an important consideration when selecting the optimal reporting frequency. Data handling costs also need to be considered and may include the software cost incurred to handle the voluminous data, the cost of human labor and computer time needed to analyze the data and the cost of additional hardware. Additional hardware may be required if the hardware that exists in the firm is not appropriate for analyzing higher frequency reports (very often due to the restrictions on memory size). The higher the reporting frequency, typically the higher will be the cost of data acquisition and handling.

In Proposition 2 we discussed how a portfolio manager may be concerned about knowing the relative rate of increase (decrease) in utility of the RMFS due to an increase (decrease) in reporting frequency. This knowledge is important to a portfolio manager who wishes to select optimal reporting frequencies by carrying out a cost-benefit analysis as shown in Proposition 4.

Proposition 4: Given that the cost of data acquisition and handling varies linearly with frequency f, the optimal reporting frequency to use will be either: (1) the maximum feasible one when the utility of the RMFS is linearly or convex increasing in f or, (2) given by the point where $\partial u/\partial f$ approaches the slope of the cost line when the utility is concave in f.

Proof: A linear cost function with respect to f is given by the following equation:

$$C(h)_{fre} = c_{fixed} + c_{variable}f \tag{22}$$

where c_{fixed} is the fixed investment in the RMFS and c_{variable} is the variable cost of acquisition and handling of data.

As we noted earlier, the value of the RMFS will be given by:

$$V(h)_{fre} = U(h)_{fre} - C(h)_{fre} \tag{23}$$

Differentiating this twice with respect to frequency f yields:

$$\frac{\partial^2 V(h)_{fre}}{\partial f^2} = \frac{\partial^2 U(h)_{fre}}{\partial f^2} - \frac{\partial^2 C(h)_{fre}}{\partial f^2} \tag{24}$$

With linear costs, $\partial^2 C(h)/\partial f^2$ equals zero and the equation reduces to:

$$\frac{\partial^2 V(h)_{fre}}{\partial f^2} = \frac{\partial^2 U(h)_{fre}}{\partial f^2} \tag{25}$$

There are four possible ways in which the linear cost and concave utility functions can behave (refer to Figure 2):

INSERT FIGURE 2 ABOUT HERE

---

(1) *The cost line does not intersect the utility curve within the region of possible reporting*

frequencies. In this case, the cost will always be more than the utility because at null reporting frequency, the RMFS will have zero utility. With cost higher than utility, the RMFS will not deliver value to the organization.

(2) *The cost line intersects the utility curve at one point*. The optimal frequency in this case will be the one where the first derivative of the utility function equals the slope $c_{variable}$ of the cost line.

(3) *The cost line touches the utility curve at one point*. In this case, the frequency corresponding to this point will be the optimal one.

(4) *The cost line intersects the utility curve at two points*. In this case, the optimal frequency will again be the one where the first derivative of the utility function equals the slope of the cost line.

When the utility curve is convex, the cost line will not be able to cut the utility function at more than one point because utility will be zero at null reporting frequency (refer to Figure 3). Under these conditions, two scenarios are possible concerning the intersection of the cost line with the utility curve:

INSERT FIGURE 3 ABOUT HERE

---

(1) *The cost line does not touch the utility curve in the zone of feasible reporting frequencies*. In this case if the cost of the RMFS is always higher than its utility, then the RMFS will not deliver value to the organization. However, if the cost of the RMFS is always less than the utility within the feasibly reporting frequency zone, then the rate of increase in value of the RMFS should increase with an increase in reporting frequency. This holds because $\partial^2V/\partial f^2$ will be greater than zero in the feasible reporting frequency zone.

(2) *The cost line either touches or intersects the utility curve at a single point corresponding to a reporting frequency f_i.* In this case, the cost of the RMFS will be more than its utility at null reporting frequency. Thus, the cost will always be more than the utility of the RMFS for frequencies less than f_i. For frequencies greater than f_i, the utility of the RMFS will be higher than its cost, and because the utility increases at a higher rate than the cost for a convex utility curve, the optimal frequency to use will the highest feasible one.

When the utility function is linear in f, two possible scenarios exist (refer to Figure 4):

INSERT FIGURE 4 ABOUT HERE

---

(1) *The cost line does not intersect the utility line.* When this occurs, the cost will always be higher than the utility because at null reporting frequency, the RMFS will have zero utility but a positive fixed cost. With cost always higher than utility, the use of the RMFS is not warranted.

(2) *The cost line intersects the utility line at a single point.* Because both cost and utility are linear and increasing in f, the slope of the utility function should be more than that of the cost function when they intersect at a point (see Figure 4). Thus, the value of the RMFS will continuously increase as frequency increases, and hence the optimal frequency to use will be the maximum feasible frequency.

We have shown that if the cost of an RMFS is linear increasing in reporting frequency, the optimal reporting frequency for the RMFS to use will be the maximum feasible frequency when the utility function is convex or linear. However, when the utility function is concave in the reporting frequency, then the optimal frequency will be the one where $\partial u/\partial f$ equals the slope of the cost line.■

Discussion: This proposition shows how the shape of the utility curve plays a role in selection of optimal data feed frequencies for risk management forecasting systems. Although, the assumption of a linear cost in f is a simplification for analysis, the analysis for cost curves of other shapes should be similar. It follows from the proposition that a portfolio manager needs to analyze utility variations with respect to changes in report frequency before selecting the optimal frequency for forecasting purposes. If the cost increases linearly in f, our results suggest that the portfolio manager should obtain reports at the highest possible frequencies when utility is linear or convex. However, when the utility is concave in f, a more careful analysis is in order. Such analysis suggests that a manager should select only those frequencies where the slope of a tangent to the utility curve most closely matches the slope of the cost line.

4.2. Response time: Individual and Joint Impacts

A delay in reporting information may seriously affect its utility: many managerial options may become inappropriate if data is received late. Although the individual impacts of response time on utility of information have been studied [EPST82], the joint impacts of response time with other data quality attributes -- report frequency, for instance -- have not been the focus of prior research efforts. A report not received on time is often worthless to managers, especially if the report frequency is high and the next report is due in a short time. On the other hand, if report frequency is low, a minor delay may not appreciably affect the expected utility of the information delivered by the report.

Our next proposition shows how the response time of data feeds interacts with their frequency to affect the utility of the risk management forecasting system. In particular, the proposition shows that the utility of high frequency reports is more affected by a delay than for low frequency reports. Before we proceed with the proposition, however, we need to define the concept of "fractional delay" in reports.

Definition 2: *Fractional delay* in reports is defined as the ratio of the time interval with delay to the time interval without delay between two successive reports.

Proposition 5: The utility of a risk management forecasting system increases at a higher rate with a decrease in response time for high frequency reports than for low frequency reports.

Proof: A delay in response time adds to the inaccuracies of predictions based on these reports. For the same time delay in two RMFSs, one producing reports at a low frequency and the other producing reports at a high frequency, the RMFS that produces reports at a higher frequency will have a greater fractional delay as compared to the RMFS that produces lower frequency reports.

To illustrate this, we let the consecutive reports for two RMFS be separated by time intervals t_{low} and t_{high}, where t_{low} is the time interval for two successive reports for the low frequency RMFS and t_{high} is for the high frequency RMFS (t_{low} should be greater than t_{high}). Let the delay in the two reports be given by Δt. Using Definition 2, the fractional delay in reports from the two RMFSs will thus be $(t_{low}+\Delta t)/t_{low}$ and $(t_{high}+\Delta t)/t_{high}$, with $(t_{low}+\Delta t)/t_{low} < (t_{high}+\Delta t)/t_{high}$.

Since the fractional delay is greater in the case when the report frequency is high, the fractional increase in inaccuracy in predictions will also be greater for higher frequency reports than for lower frequency reports for same amount of delay. Going a step further, we can also say that the increase in predictive inaccuracy, given by Δs, for a unit increase in response time (r) of the reports will be higher for high frequency reports than for low frequency reports. Hence:

$$\frac{\left|\frac{\partial \Delta s}{\partial r}\right|_{f=f_{high}} - \left|\frac{\partial \Delta s}{\partial r}\right|_{f=f_{low}}}{f_{high} - f_{low}} \geq 0. \tag{26}$$

Now, to prove Proposition 5 we need to show that the decrease in utility of the RMFS with an increase in the response time is higher for high frequency reports than for low frequency reports. In equational form, we need to show that $\partial^2 T/\partial f \partial r \leq 0$.

Differentiating $\partial T/\partial f$ with respect to r yields:

$$\frac{\partial^2 T}{\partial f \partial r} = \frac{\partial^2 u}{\partial f \partial r} p(\Delta s) p(R) + \frac{\partial u}{\partial r}\frac{\partial p(\Delta s)}{\partial f} p(R) + \frac{\partial u}{\partial r} p(\Delta s)\frac{\partial p(R)}{\partial f} + u\frac{\partial^2 p(\Delta s)}{\partial f \partial r} p(R) + u\frac{\partial p(R)}{\partial f}\frac{\partial P(\Delta s)}{\partial r}$$
$$+\frac{\partial u}{\partial f}\frac{\partial p(\Delta s)}{\partial r} p(R) + u\frac{\partial^2 p(R)}{\partial f \partial r} p(\Delta s) + u\frac{\partial p(R)}{\partial r}\frac{\partial p(\Delta s)}{\partial f} + \frac{\partial u}{\partial f}\frac{\partial p(R)}{\partial r} p(\Delta s) \quad (27)$$

where $p(\Delta s)$ is used as an abbreviation for $p(\Delta s \mid R, m_h)$.

Since the distribution of the data set is assumed to be invariant with respect to small changes in frequency and response time, all terms in (27) involving variations in p(R) should be zero.

Differentiating u and $\partial u/\partial r$ in (27) with respect to f results in:

$$\frac{\partial u}{\partial r} = \frac{\partial u}{\partial \eta}\frac{\partial \eta}{\partial(\Delta s)}\frac{\partial(\Delta s)}{\partial r} \quad (28)$$

and,

$$\frac{\partial^2 u}{\partial f \partial r} = \frac{\partial u}{\partial \eta}\frac{\partial \eta}{\partial(\Delta s)}\frac{\partial^2(\Delta s)}{\partial f \partial r} + \frac{\partial u}{\partial \eta}\frac{\partial^2 \eta}{\partial f \partial(\Delta s)}\frac{\partial(\Delta s)}{\partial r} + \frac{\partial^2 u}{\partial f \partial \eta}\frac{\partial \eta}{\partial(\Delta s)}\frac{\partial(\Delta s)}{\partial r} \quad (29)$$

Because we assumed that utility was linear in hedge efficiency, we know that $\partial u/\partial \eta \geq 0$. According to our definition of a highly responsive hedging scheme, $\partial \eta/\partial(\Delta s) \leq 0$. And because with higher response time the inaccuracy in predictions from the system is expected to increase, we also should have $\partial(\Delta s)/\partial r \geq 0$. This enables us to conclude that $\partial u/\partial r \leq 0$.

With hedge efficiency, η, assumed to be linear in Δs for a highly responsive hedging scheme, the second term in (29) equals zero. And, because we assumed that utility varies linearly in η, the third term in (29) also should equal zero. Taking these results into consideration and combining them with already derived results that $\partial^2 \Delta s/\partial f \partial r \geq 0$ (see Equation 26), $\partial u/\partial \eta \geq 0$ and $\partial \eta/\partial(\Delta s) \leq 0$, we get $\partial^2 u/\partial f \partial r \leq 0$ from (29). This makes the first term in (27) less than or equal to zero. Because the reliability of the forecasting model increases with an increase in frequency, $\partial p(\Delta s)/\partial f$ is less than or equal to zero. This makes the second term in (27) less than or equal to zero.

The reliability of the forecasting model should reduce with an increase in response time, thus making the sixth term of (27) less than or equal to zero. However, this reduction should be higher for lower frequency than for higher frequency as the number of observations used for forecasting is less in the former than in the latter case. This means that the fourth term, $\partial^2 p(\Delta s)/\partial f \partial r$, will be less than zero.

Finally, because the terms involving variations in p(R) are all equal to zero, we deduce that $\partial^2 T/\partial f \partial r \leq 0$. This completes the proof.■

Discussion: Proposition 5 shows how the impact of report frequency on the utility of an RMFS varies with the delay in reports. The proposition states and proves that the utility of a system that uses high frequency data feeds is more negatively affected than the utility of a system that uses lower frequencies when reports are received late.

Example: In our MBS example, the decline in predictive accuracy given by $\partial(\Delta s)/\partial r$ is greater for high frequency reports than for low frequency reports as can be seen below.

$$\frac{\partial(\Delta s)}{\partial r} = 2f^2 r^{-3} \varepsilon^{-2}. \tag{30}$$

As expected, the rate of increase in utility with a decrease in response time r declines with an increase in frequency as shown by the expression for $\partial^2 T/\partial f \partial r$:

$$\frac{\partial^2 T}{\partial f \partial r} = -4k_2(1-\Delta s_{\max}) f r^{-3} \varepsilon^{-2} - 16 k_2 f^3 r^{-5} \varepsilon^{-4}. \tag{31}$$

Given our earlier assumption that the value of k_2 is greater than or equal to zero, $\partial^2 T/\partial f \partial r$ is less than or equal to zero.

4.3. Accuracy and Value of Data Feeds

As the accuracy of data increases, the utility of a risk management forecasting system also is expected to increase. However, a more interesting research concern is the interaction between accuracy and report frequency.

Proposition 6: The utility of a risk management forecasting system increases at a lower rate with an increase in report frequency for higher amount of report inaccuracies than for lower amount.

Proof: If the inaccuracy in reports is high, the effect of an increase in frequency on the predictive accuracy of the model gets diluted. Hence,

$$\frac{\left|\frac{\partial \Delta s}{\partial f}\right|_{\varepsilon=\varepsilon_{high}} - \left|\frac{\partial \Delta s}{\partial f}\right|_{\varepsilon=\varepsilon_{low}}}{\varepsilon_{high} - \varepsilon_{low}} \geq 0. \tag{32}$$

Now, to prove proposition 6, we need to show that the increase in utility of the RMFS with an increase in frequency is higher for lower amount of report inaccuracies than for higher amount of report inaccuracies. In equational form, we need to show that $\partial^2 T/\partial \varepsilon \partial f \leq 0$. Differentiating $\partial T/\partial f$ with respect to ε, we get:

$$\frac{\partial^2 T}{\partial\varepsilon\partial f} = \frac{\partial^2 u}{\partial\varepsilon\partial f}p(\Delta s)p(R) + \frac{\partial u}{\partial f}\frac{\partial p(\Delta s)}{\partial\varepsilon}p(R) + \frac{\partial u}{\partial f}p(\Delta s)\frac{\partial p(R)}{\partial\varepsilon} + u\frac{\partial^2 p(\Delta s)}{\partial\varepsilon\partial f}p(R) + u\frac{\partial p(R)}{\partial\varepsilon}\frac{\partial P(\Delta s)}{\partial f} \qquad (33)$$
$$+ \frac{\partial u}{\partial\varepsilon}\frac{\partial p(\Delta s)}{\partial f}p(R) + u\frac{\partial^2 p(R)}{\partial\varepsilon\partial f}p(\Delta s) + u\frac{\partial p(R)}{\partial f}\frac{\partial p(\Delta s)}{\partial\varepsilon} + \frac{\partial u}{\partial\varepsilon}\frac{\partial p(R)}{\partial f}p(\Delta s).$$

The first term in equation involves $\partial^2 u/\partial\varepsilon\partial f$, which is given by the following expression:

$$\frac{\partial^2 u}{\partial\varepsilon\partial f} = \frac{\partial u}{\partial\eta}\frac{\partial\eta}{\partial\Delta s}\frac{\partial^2\Delta s}{\partial\varepsilon\partial f} + \frac{\partial u}{\partial\eta}\frac{\partial^2\eta}{\partial\varepsilon\partial\Delta s}\frac{\partial\Delta s}{\partial r} + \frac{\partial^2 u}{\partial\varepsilon\partial\eta}\frac{\partial\eta}{\partial\Delta s}\frac{\partial\Delta s}{\partial f} \qquad (34)$$

Following our assumptions (see proof for Proposition 5) and taking into account the results from (32), we have $\partial^2 u/\partial\varepsilon\partial f \leq 0$. This makes the first term in (33) less than or equal to zero. Since the reliability of the model decreases with an increase in inaccuracy, $\partial p(\Delta s)/\partial\varepsilon$ is less than or equal to zero. This makes the second term in (33) also less than or equal to zero.

The variation in utility u with an increase in inaccuracy of reports is given by:

$$\frac{\partial u}{\partial\varepsilon} = \frac{\partial u}{\partial\eta}\frac{\partial\eta}{\partial(\Delta s)}\frac{\partial(\Delta s)}{\partial\varepsilon} \qquad (35)$$

The predictive accuracy is expected to decrease with an increase in inaccuracy in data, and so $\partial u/\partial\varepsilon$ should be less than or equal to zero. The reliability of the model should increase with an increase in frequency, thus making the sixth term of (33) less than or equal to zero. However, this increase should be lower for inaccurate data than for accurate data. This implies that $\partial^2 p(\Delta s)/\partial\varepsilon\partial f \leq 0$. This makes the fourth term less than or equal to zero also. Since the terms involving variations in p(R) are all equal to zero, we deduce that $\partial^2 T/\partial\varepsilon\partial f \leq 0$.■

Discussion: Proposition 6 shows that if the data are inaccurate, the payoffs a portfolio manager receives when using a high frequency RMFS are not going to be cost efficient.

Thus, if a portfolio manager is contemplating investing in improving the data quality of reports, he should first attempt to improve the data accuracy before thinking about improvements in report frequencies.

Example: For our MBS example, the increase in predictive accuracy due to an increase in report frequency as given by $\partial(\Delta s)/\partial f$ is lower for more inaccurate reports than for less inaccurate reports as can be seen below.

$$\frac{\partial(\Delta s)}{\partial f} = -2fr^{-2}\varepsilon^{-2}. \tag{36}$$

Note that the rate of increase in utility associated with an increase in frequency f declines with data inaccuracies increase, as shown by the expression for $\partial^2 T/\partial\varepsilon\partial f$:

$$\frac{\partial^2 T}{\partial\varepsilon\partial f} = -4k_2(1-\Delta s_{max})fr^{-2}\varepsilon^{-3} - 16k_2 f^3 r^{-4}\varepsilon^{-5}. \tag{37}$$

which, given our assumptions, is less than or equal to zero.

5. CONCLUSIONS

In the preceding sections, three attributes of data quality, including frequency, response time, and accuracy were studied in detail. Their impact on the utility of a risk management forecasting system was studied in terms of the results from six propositions. While previous studies in literature mainly investigated individual impacts of data quality attributes on the value of information, this research focused on both individual and joint impacts of data quality attributes.

A number of interesting results were obtained. We showed that the utility of a risk management forecasting system (RMFS) increased with the frequency of reports, but the rate of this increase may vary from situation to situation. Therefore, to determine optimal reporting frequencies, a portfolio manager would need to carefully study the characteristics of the utility curve for the RMFS. Also, the magnitude of increase in utility with an increase in reporting frequency depended on the responsiveness of a hedging scheme that maps predicted states to hedge efficiencies. We also saw that an increase in response time for reports decreased the utility of an RMFS, but this decrease was much higher for high frequency reports than for low frequency reports. Accuracy was also shown to interact with frequency and we found that inaccuracy diminished the increase in utility of an RMFS when reporting frequency increases.

The applicability of this research is diverse. Although the discussion in the paper was largely centered around financial risk management forecasting systems, the results are easily extended to other non-financial areas, if the assumptions that we discussed hold. A few examples are: collection of suitable data in the domains of marketing, selection of variables to be considered for quality control in manufacturing and selection of appropriate decision variables in auditing. However, one of the major limitations of this work is that the results are dependent on the assumptions of the model. It would be interesting to carry out a sensitivity analysis of these results by relaxing some of these assumptions.

This paper provides a theoretical foundation to investigate the relationship between the different attributes of data quality and value of risk management data feeds. For this study we limited our attention to the data feed component of risk management technology. We plan to extend this approach by considering the effects of data quality on value of risk management systems taking into account the inter-model variations in risk management.

References

[AHIT89] Ahituv, N. "Assessing the Value of Information: Problems and Approaches." In *Proceedings of the Tenth International Conference on Information Systems*, Boston, Massachusetts, 1989, pp. 315-325.

[AREN89] Arend, M. "Financial Databases Flag Overseas Opportunities," *Wall Street Computer Review*, October 1989, pp. 24-70.

[BAGM75] Bagman, S., "The Evaluation and Choice of Internal Information Systems Within a Multiperson World," *Journal of Accounting Research*, Spring 1975, pp. 1-15.

[BALL85] Ballou, D. P., and Pazer, H. L. "Modeling Data and Process Quality in Multi-Input, Multi-Output Information Systems," *Management Science*, Volume 31, Number 2, February 1985, pp. 152-162.

[BALL89] Ballou, D. P. and Tayi, G. K. "Methodology for Allocating Resources for Data Quality Enhancement," *Communications of the ACM*, Volume 32, Number 3, pp. 320-329, March 1989.

[BARU89] Barua, A.; Kriebel, C. H; and Mukhopadhyay, T. "MIS and Information Economics: Augmenting Rich Description with Analytical Rigor in Information Systems Design." In *Proceedings of the Tenth International Conference on Information Systems*, Boston, Massachusetts, 1989, pp. 327-339.

[BLAC53] Blackwell, D. "Equivalent Comparison of Experiment." *Annual Mathematical Statistics*, June 1953, pp. 265-272.

[CUSH74] Cushing, B. E., "A Mathematic Approach to the Analysis and Design of Internal Control Systems," *The Accounting Review*, Volume 49, Number 1, January 1974, pp. 24-41.

[DEMS85] Demski, J. S. *Information Analysis*, Second Edition, Addison-Wesley Publishing, Reading, MA, 1985.

[DOHE85] Doherty, N. A. *Corporate Risk Management*, McGraw-Hill, Inc., New York, NY, 1985.

[EPST82] Epstein, B. J.; and King, W. R., "An Experimental Study of the Value of Information," *Omega*, Volume 10, Number 3, pp. 249-258, July 1982.

[GROS80] Grossman, S. J.; and Stiglitz, J. E., "On the Impossibility of Informationally Efficient Markets," *The American Economic Review*, Volume 70, Number 3, June 1980, pp. 393-408.

[HAYR89] Hayre, L. S.; Lauterbach, K. and Mohebbi C., "Prepayment Models and Methodologies." *In Advances and Innovations in the Bond and Mortgage Markets*, Frank J. Fabozzi (editor), Probus Publishing Company, Chicago, IL, 1989.

[HILT79] Hilton, R. W. "The Determinants of Cost Information Value: An illustrative Analysis." *Journal of Accounting Reviews*, Volume 17, Number 2, Autumn 1979, pp. 411-435.

[HILT81] Hilton, R. W. "The Determinants of Information Value: Synthesizing Some General Results." *Management Science*, Volume 27, Number 1, January 1981, pp. 57-64.

[IJIR73] Ijiri, Y.; and Itami, H. "Quadratic Cost-Volume Relationship and the Timing of Demand Information," *The Accounting Review*, Volume 48, Number 3, October 1973, pp. 724-737.

[KIHL74] Kihlstrom, R., "A General Theory of Demand for Information About Product Quality," *Journal of Economic Theory*, August 1974, pp. 413-439.

[MARS72] Marschak, J., and Radner, R. *Economic Theory of Teams*, Yale University Press, New Haven, CT, 1972.

[McCA85] McCardle, K. F., "Information Acquisition and the Adoption of New Technology," *Management Science*, Volume 31, Number 11, November 1985, pp. 1372-89.

[PINK87] Pinkus, S. M.; Hunter, S. M.; and Roll, R., "An Introduction to the Market and Mortgage Analysis," In *Advances in Bond Analysis & Portfolio Strategies*, Frank J. Fabozzi and T. Dessa Garlicki (editors), Probus Publishing Company, Chicago, IL, 1987.

[SCHM90A] Schmerken, I. "Software to Ride Out Interest-Rate Turbulence," *Wall Street Computer Review*, March 1990, pp. 14-76.

[SCHM90B] Schmerken, I. "Wall Street Struggles to Balance Global Risk," *Wall Street Computer Review*," August 1990, pp. 12-56.

[SHAL89] Shale, T. "The Great Risk-Management Systems Failure," *Euromoney*, February 1989.

[STEV87] Stevenson, M. "International Banking a Game of Skill As Well," *The*

[IJIR73] Ijiri, Y.; and Itami, H. "Quadratic Cost-Volume Relationship and the Timing of Demand Information," *The Accounting Review*, Volume 48, Number 3, October 1973, pp. 724-737.

[KIHL74] Kihlstrom, R., "A General Theory of Demand for Information About Product Quality," *Journal of Economic Theory*, August 1974, pp. 413-439.

[MARS72] Marschak, J., and Radner, R. *Economic Theory of Teams*, Yale University Press, New Haven, CT, 1972.

[McCA85] McCardle, K. F., "Information Acquisition and the Adoption of New Technology," *Management Science*, Volume 31, Number 11, November 1985, pp. 1372-89.

[PINK87] Pinkus, S. M.; Hunter, S. M.; and Roll, R., "An Introduction to the Market and Mortgage Analysis," In *Advances in Bond Analysis & Portfolio Strategies*, Frank J. Fabozzi and T. Dessa Garlicki (editors), Probus Publishing Company, Chicago, IL, 1987.

[SCHM90A] Schmerken, I. "Software to Ride Out Interest-Rate Turbulence," *Wall Street Computer Review*, March 1990, pp. 14-76.

[SCHM90B] Schmerken, I. "Wall Street Struggles to Balance Global Risk," *Wall Street Computer Review*," August 1990, pp. 12-56.

[SHAL89] Shale, T. "The Great Risk-Management Systems Failure," *Euromoney*, February 1989.

[STEV87] Stevenson, M. "International Banking a Game of Skill As Well," *The*

Economist, March 21 1987.

[WEST86] Weston, J. F. and Copeland, T. E. *Managerial Finance*, The Dryden Press, New York, NY, 1986.

[WILS75] Wilson, R. "Informational Economies of Scale," *Bell Journal of Economics*, 1975, pp. 184-195.

Figure 1: Previous Studies on Impacts of Data Quality Variations

DATA QUALITY ATTRIBUTE	STUDY	CONTEXT OF STUDY	RELEVANT CONTRIBUTIONS	COMMENTS
FREQUENCY	Barua, Kriebel and Mukhopadhyay [BARU89]	Theoretical with manufacturing examples	Increase in reporting frequency does not imply an increase in utility.	For frequencies that are integer multiplications of each other, the results may not hold; joint effects ignored.
	Cushing [CUSH74]	Internal control systems	Studied propagation of errors due to imperfect control timing.	Utility and economic value not considered explicitly; joint impacts of transaction size, process probabilities and timing of control not investigated.
ACCURACY	Wilson [WILS75]	Machine setting and timing of deliveries	Information value increasing in accuracy.	Accuracy measured by cell length in partition; cost not considered.
	Ijiri and Itami [IJIR73]	Quadratic cost-volume decision with uncertain demand	Information value increasing in accuracy.	Accuracy is measured by expected variance in state posterior given a signal. No joint impacts of other attributes considered.
COST	Marschak and Radner [MARS72]	Market speculation	Cost can be differentiated into cost of transmission, acquiring and producing information. Latter two affect the value.	Cost of transmission is considered invariant for all practical reasons for different data feeds.
	Ballou and Pazer [BALL85]	Accounting systems	Utility can be equated to an opportunity cost by identifying losses that would result by not having high enough quality.	Cost of opportunity can be added to the cost of acquisition for data quality evaluation. Joint impacts of different data quality attributes considered.
RESPONSE TIME	Barua, Kriebel and Mukhopadhyay [BARU89]	Manufacturing	Earlier response time will be preferred if the maximum value of action decreases with time.	Joint impacts with frequency not considered. For high frequency if a report is delayed it effectively reduces the frequency.

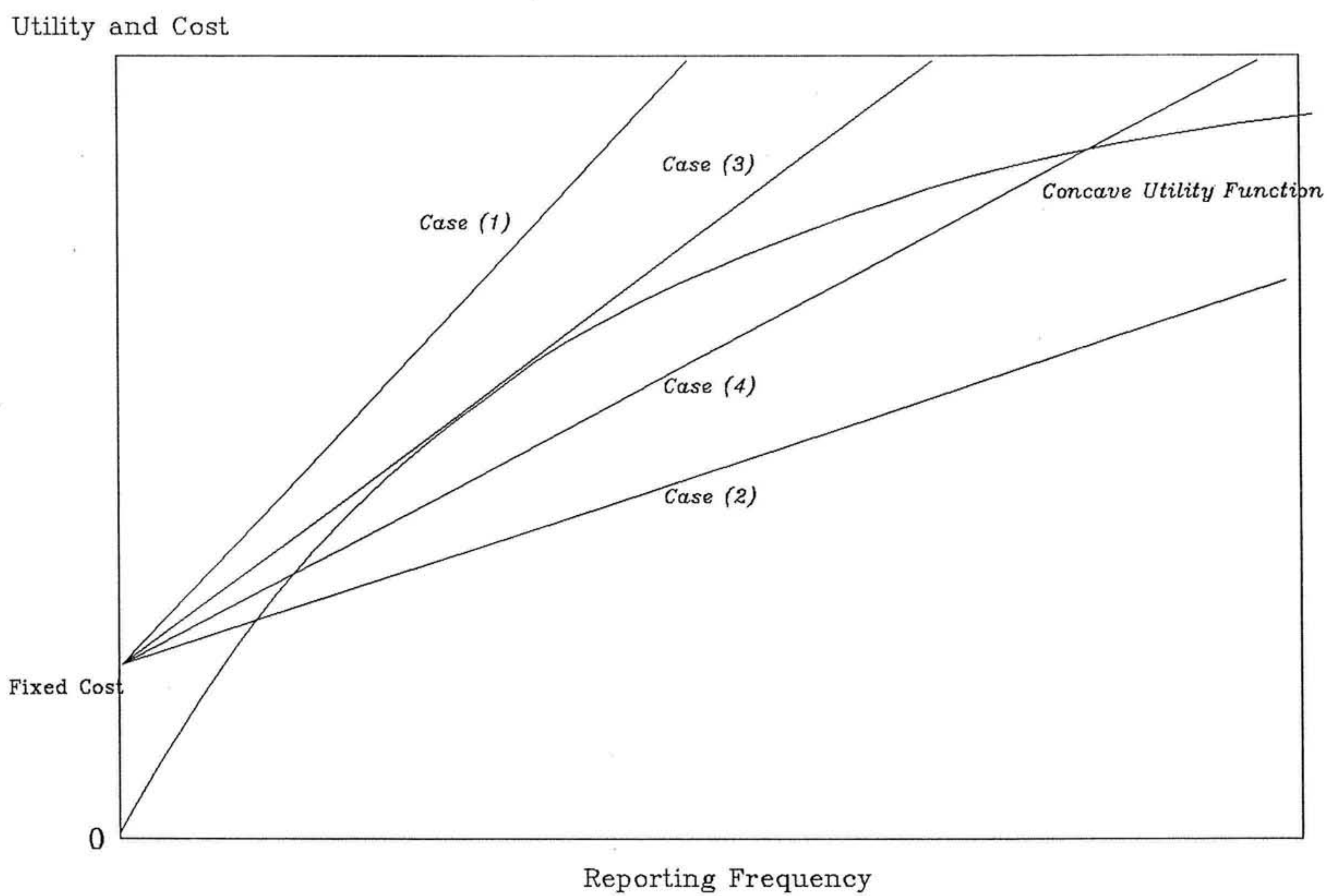

Figure 2: Concave Utility and *Linear Costs*

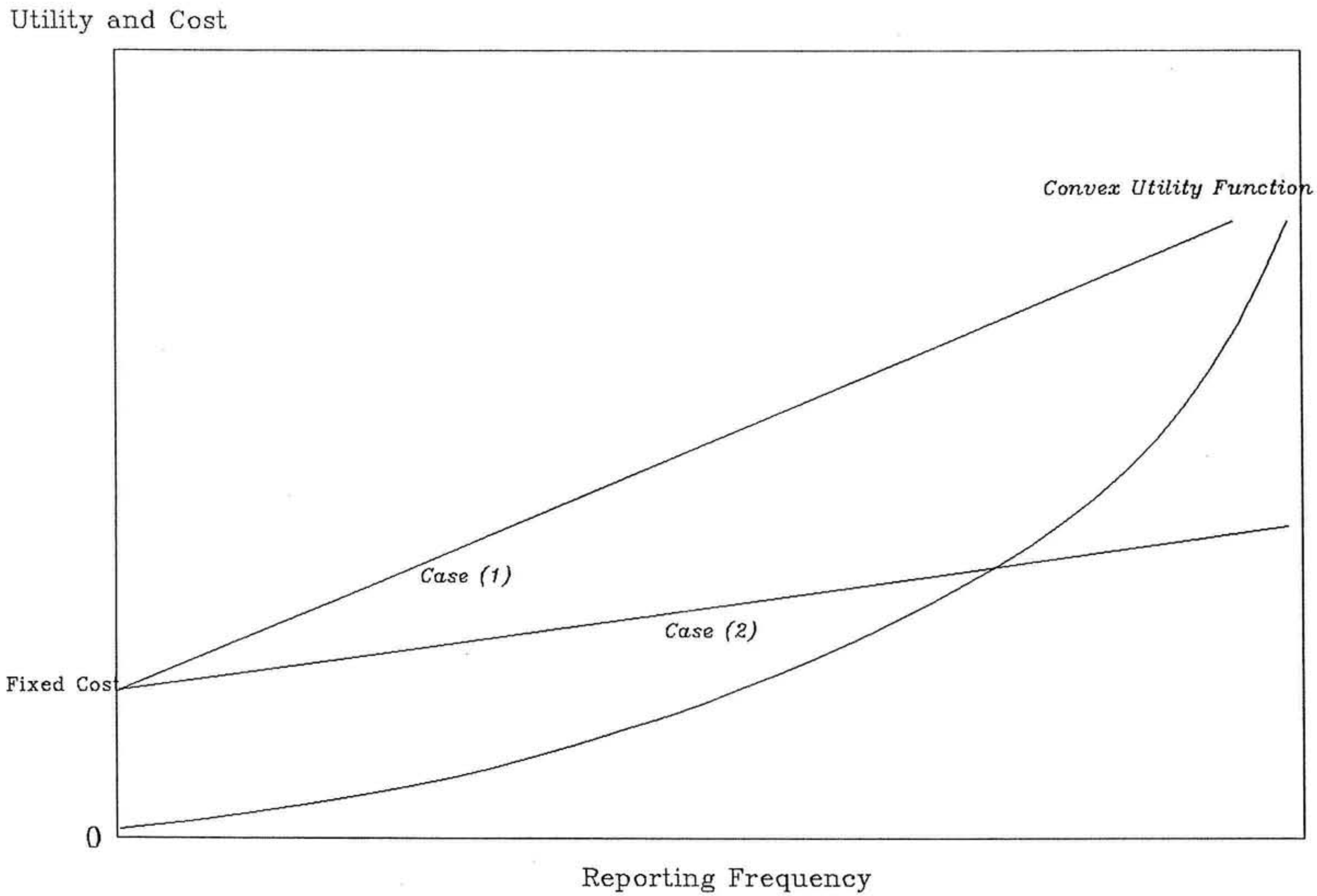

Figure 3: Convex Utility and Linear Costs

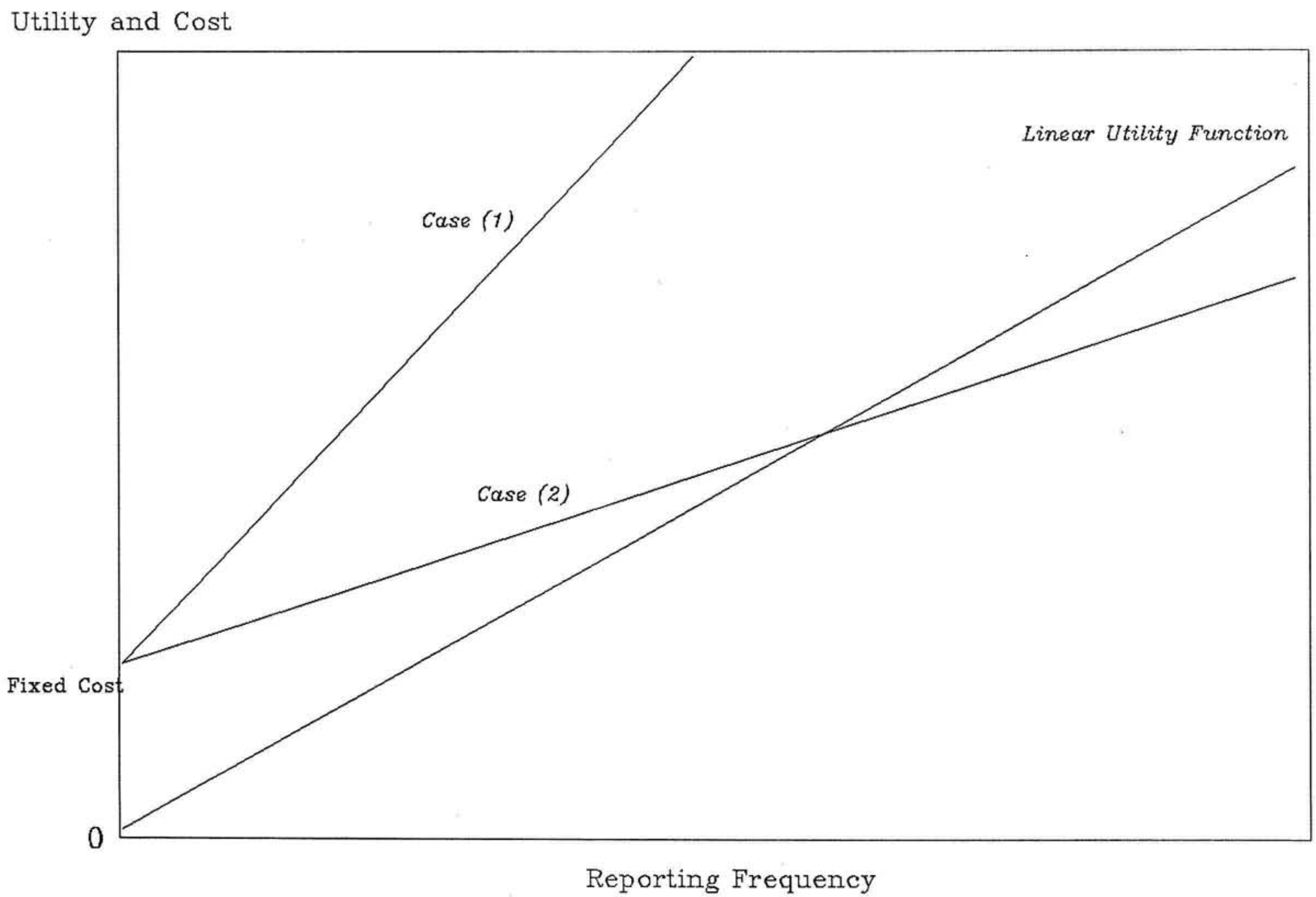

Figure 4: Linear Utility and Linear Costs